AD-A096 965 FLORIDA UNIV GAINESVILLE CENTER FOR MATHEMATICAL SYS--ETC F/G 5/3
IDENTIFIABILITY AND PROBLEMS OF MODEL SELECTION FOR TIME-SERIES--ETC(U)
1980 R E KALMAN AFOSR-76-3034
UNCLASSIFIED AFOSR-TR-81-0229 NL

1 OF 1
AD
A096965

END
DATE
FILMED
4-81
DTIC

AD A096965

LEVEL II

(2)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (19) (18) AFOSR-TR- 81-0229 ✓	2. GOVT ACCESSION NO. AD-A096
	3. RECIPIENT'S CATALOG NUMBER 965
4. ~~TITLE (and Subtitle)~~ (6) IDENTIFIABILITY AND PROBLEMS OF MODEL SELECTION FOR TIME-SERIES ANALYSIS IN ECONOMETRICS	5. TYPE OF REPORT & PERIOD COVERED (9) Interim technical rep
	6. ~~PERFORMING ORG. REPORT NUMBER~~
7. AUTHOR(s) (10) R. E. Kalman	8. CONTRACT OR GRANT NUMBER(s) AFOSR-76-3034
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Mathematical System Theory University of Florida Gainesville, FL 32611 411710	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) 2304/A6 (17) A6 61102F
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research /NM Bolling AFB, DC 20332	12. REPORT DATE (11) 1980 (12) 38
	13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (15) AFOSR-76-3034, DAAG29-77-G-0225	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release;
distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

A

18. SUPPLEMENTARY NOTES

Presented as an invited lecture at the Fourth World Congress of the Econometric Society, Aix-en-Provence, France, August 30, 1980

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Identification, simultaneous model selection

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Identification and model selection in econometrics are examined from a system theoretic point of view. A critical study of the existing techniques in econometrics is carried out through several examples from literature.

DTIC FILE COPY

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 408710

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AFOSR-TR- 81-0229

Presented as an invited lecture at the Fourth World Congress of the Econometric Society, Aix-en-Provence, FRANCE, August 30, 1980.

IDENTIFIABILITY AND PROBLEMS OF MODEL SELECTION FOR TIME-SERIES ANALYSIS IN ECONOMETRICS*

by

R. E. Kalman

Swiss Federal Institute of Technology, Zürich, SWITZERLAND
University of Florida, Gainesville, FL USA

DTIC
SELECTE
MAR 30 1981
S D
A

---

* This research was supported in part under US Air Force Grant AFOSR-76-3034 and US Army Research Grant DAA 29-77-G-0225 through the Center for Mathematical System Theory, University of Florida, Gainesville, FL 32611 USA.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification______
By______
Distribution/
Availability Codes
Dist | Avail and/or Special
A

Approved for public release;
distribution unlimited.

81 3 27 166

1. BACKGROUND AND PERSPECTIVE

Econometrics was born in the 1920's from the hope that it would be the embodiment of a noble dream. Today, with the hindsights of more than fifty years, this dream may be described precisely, as follows:

> "Economics deals with complex phenomena which make it impossible to isolate quantitative relationships between important variables (say, taxations vs. savings). Nor is it possible to perform experiments or direct observations that would isolate these relationships or at least diminish the noise level under which they are observable. We possess, however, innumerable time series generated by the economic "forces" we wish to discover. By constructing models for economic time series we may hope to obtain indirect access to the desired quantitative economic relationships since these are "encoded" in the models and therefore must be recoverable from the models' structure and parameters. Moreover, since economic truths are immutable, at least in the short run, there is no reason, in principle, why the models cannot be accurately determined, in spite of disturbances, errors, irrationalities, expectations, and other random influences which afflict the available time series."

My personal analysis of the preceding scenario is that it forces the study of economics to be a system-theoretic endeavor. Economics as a science requires methods which are effective in the exploration and explanation of interacting phenomena. Thus implementing the dream of econometrics becomes a problem in system theory, a discipline which was not even in sight in the 1920's.

But, unfortunately, the actual evolution of econometrics took a different route and came to be dominated by statistics; the result is that today econometrics has drifted apart from the scientific mainstream of the 1950's and 1960's.

During this period system theory has reached a certain level of maturity. It is now beginning to provide the scientific framework within which basic ideas of economics and econometrics can be reexamined and subjected to deep analysis and critique. Some of its current beliefs, procedures, and even results must be changed if econometrics is to continue to play its role on a world scene where massive computer analysis and advanced mathematical methods of modeling compete with it in attacking the same basic problems.

As I have pointed out in several previous publications (see KALMAN [1979a, 1979b, 1980a, 1980b]), research in modeling cannot be done meaningfully if it is viewed as a problem which is inseparable from a particular field of application. Habits of economic modeling cannot be justified solely by economic reasoning. Modeling has its own logic, independently of what is being modeled. And the questions raised by modeling logic may be far more important than practical problems such as reliability of data, use of prior economic theory, statistical methodology, and the like.

The papers just quoted are concerned with general, even philosophical, questions of modeling. In this paper we shall focus on a specific question: the concept of identifiability.

Econometricians have become accustomed to formalize the process of modeling by asking whether or not the parameters describing a model may be deduced as a function of (i. e., uniquely determined by) the "information" or "data" which is available for the construction of the model. If this requirement is met, the parameters are said to be identifiable.

Such a notion of identifiability is unexceptional on purely conceptual grounds. But this is not enough. The concept of "parameter identifiability" must also make system-theoretic sense. Whether this is so or not can be objectively decided because the notion of a model is a precisely defined concept in system theory.

When "identifiability" is critically examined, several important new questions arise which have not been considered by econometricians.

First, it is nontrivial whether or not the model is well-defined in the mathematical sense. There are many instances in the econometrics literature where this is just not the case. System theory provides criteria for discarding models which are not well-defined.

Second, assuming now that the model is indeed well-defined, it follows, essentially as a matter of definition, that the model or models compatible with the data is or are necessarily identifiable in the abstract sense. The difficulty is nonuniqueness, not identifiability.

Third, there is the question of parametrization of models. In econometrics, "parameters" are used in the descriptive sense; in other words, simply to give a mathematical specification of the model. The trouble is that a model is an abstract mathematical object. It is subject to various assumptions and restrictions in addition to its describing parameters. Therefore its intrinsic parametrization, which reflects its precise mathematical attributes, is usually quite different from the naive (descriptive) parametrization. The problem of parametrization of a model, as of any abstract mathematical object, is a highly nontrivial mathematical problem and is not accessible to the kind of elementary reasoning that has been used in econometrics or time-series analysis. The confusion between descriptive and intrinsic parameters is deep-seated and it is not easy to clarify it on the intuitive level.

In short, "parameter identifiability" as a scientific concept is of no utility. Conceptually, it is just not a good tool for probing deeper properties of models. The basic tasks are to study models which arise in time-series identification problems and to do the concomitant mathematical research concerning their invariants, local coordinates, or intrinsic parameters---all three terms being synonymous.

The issues involved will be illustrated in Sections 3 through 5 by simple examples taken from the econometric literature. We shall stress the conceptual rather than the technical aspects. A critical and mathematical treatment, drawing upon a wider range of the econometric literature, will be given in KALMAN [1981].

The contribution of this paper, it must be emphasized, does not reside in proposing a new methodology for econometrics. Rather, the contribution consists in pointing out serious intrinsic limitations of the existing methodological state of affairs in the light of superior knowledge already available in system theory. The required changes in econometrics are fundamental, important, and not just a symptom of disagreement between researchers from different fields or having different objectives.

Some may find these changes not to their liking. I am sorry but as a scientist it is my duty to call attention to the reasons which mandate change. If econometricians ignore the very real difficulties discussed here, their discipline will become sterile and effete, like a graying man clinging to a dream, perhaps even a love, of his early youth, still hoping he can continue to work as he always did, without taking notice of the many wonderful things that were born into the world after the conception of his dream.

2. THE MODEL AS A LINEAR DYNAMICAL SYSTEM

Econometric modeling from time series requires system theory at its very first step: *the definition of the model*. This may seem a triviality but turns out to be crucial.

Most of the time-series literature is concerned only with *linear* models. Therefore we shall restrict our attention (here) to the corresponding concept of a *linear system*. This concept has been axiomatized (see KALMAN, FALB, and ARBIB [1969, Chapters 1, 2, and 10]) and provides a reference point to which all further definitions and results can be compared. It also provides a means whereby the current status and claims of time-series modeling may be scientifically assessed, which is what is done in KALMAN [1981].

In the definition of a system, precise technical meaning must be given to the attributes *linear*, *finite* (-dimensional and finitely parametrized), *multi-input/multi-output*, *constant* (= time-independent in its structural properties), and *dynamical*. These words are all incorporated in the standard definition* which comes in two versions. For continuous-time, that is, with the time set $T = \mathbb{R}$ = real numbers, a system Σ is defined by

$$(2.1) \qquad \frac{dx}{dt} = Fx + Gu(t), \quad y(t) = Hx(t), \quad T \in \mathbb{R};$$

for discrete-time, that is, with the time set $T = \mathbb{Z}$ = integers, a system Σ is given by

---

* The notations in (2.1) and (2.2), which I introduced around 1960 to honor my great teacher, F. G. H. Linear, have been universally adopted.

(2.2) $x(t + 1) = Fx(t) + Gu(t), \quad y(t) = Hx(t), \quad t \in \underline{\underline{Z}}.$

In (2.1-2.2), the real (or complex) vectors x, u, and y are called state, input, and output, respectively; F, G, H are matrices with constant real (or complex) coefficients.

It is rather obvious from its definition that a "system" (which will be our shorthand for the precise terminology concomitant with (2.1-2)) is really defined by the "data" (F, G, H). So we frequently write simply $\Sigma = (F, G, H)$. The notations have been intentionally selected in such a way that there is no built-in distinction between the continuous-time and discrete-time cases. Most of the system-theoretic questions are purely algebraic in nature, based on properties of F, G, and H, and therefore such a distinction is not necessary; the results hold simultaneously for continuous-time and discrete-time.

The concept of a "system" goes back to Newtonian mechanics, with the very important addition of the concepts of "inputs" and "outputs". A good mental model, especially in discrete time, is a computer. The formulation of the basic definitions is conceptually valid in complete generality, without the assumption of linearity (see KALMAN, FALB, and ARBIB, [1969, Chapter 1]). But linearity becomes essential if universal (that is, system-theoretic) mathematical results are wanted, not just definitions. The power of mathematics, as currently applied to system theory, stems almost entirely from the word "linear".

Definitions (2.1-2) formalize the concept of a system in the axiomatic style. They provide a highly convenient starting point for further discussions. Sometimes (2.1-2) are called the internal definitions of a system, alluding to the fact that the definition is stated in terms of the internal or state variables (the components of the vector x).

To relate the axiomatics to the real world, it is necessary to introduce the word "behavior". It has about the same meaning in system theory as in econometrics or other applied fields. "Behavior" means directly observable properties of a system; for a deterministic system "observation" means, by definition (!), knowledge of the input and the output; the state is always to be regarded as nonobservable. (In stochastic system theory, observation of the inputs is replaced by the postulated knowledge of their probability distribution. For us here, this distinction is a side issue.)

The output of a linear system must be linearly (causally) related to its input. Using nothing more than the mathematical definition of linearity, this implies that for a discrete-time system the input/output relation must take the form

$$(2.3) \quad y(t) = \sum_{\tau=0}^{t-1} A_{t-\tau} u(\tau), \quad x(0) = 0.$$

We may view (2.3) as the external definition of a system. Thus "behavior" for a linear discrete-time system is quantified by the specification of an infinite sequence of matrices

$$(2.4) \quad S = (A_1, A_2, \dots).$$

For continuous-time systems, the definition of S is again the same as (2.4). (However, (2.3) must be replaced by the convolution integral and the derivation of (2.4), which we omit here, turns out to involve nontrivial mathematical technicalities.)

Note that (2.4) represents deterministic (nonstochastic) behavior.

The behavior of a system given by (2.1-2) is easily calculated by the algebraic formulas

$$(2.5) \quad A_t := HF^{t-1}G, \quad t = 1, 2, \dots .$$

Whether we take the internal or the external point of view in defining a system, we must always bear in mind that we are dealing with an abstract mathematical object. This fact must not be obscured by the fact that to describe either (F, G, H) or S we make use of numbers (or, according to common usage, "parameters"), namely the elements of the matrices (F, G, H) and of the (infinitely many) matrices $A_1, A_2, \dots$.

The descriptive parameters of a system are not at all the same as its intrinsic parameters. The reason is that a "system" is really a nonlinear object and does not admit parameters in the same simple sense as, e. g., vectors in the space $\underline{\underline{R}}^n$. To (intrinsically) parametrize (or coordinatize) a family of systems means that each system in the family is given by a unique set of numbers which is not at all a requirement for a descriptive parametrization.

To understand this point better, let us emphasize that the definitions (2.1-2) are subject to an equivalence relation which arises from regarding two systems as essentially the same when their (external) behavior is the same. It is easy to see that a change of coordinates in the state space $X = R^n$, which is written as $\Sigma \to \hat{\Sigma}$ and defined by the relations

$$(2.6) \quad \begin{cases} \hat{F}T = TF, \\ \hat{G} = TG \\ \hat{H}T = H, \end{cases} \qquad (\det T \neq 0),$$

implies $S_\Sigma = S_{\hat{\Sigma}}$, that is, preservation of behavior. The converse is also true, provided we relax the condition $\det T \neq 0$.

To (intrinsically) parametrize a family of systems Σ subject to this equivalence relation is a nontrivial problem. Unless this is done, Σ cannot be identified from S because the identification problem by definition, relates only to the behavior of a system, and not to its axiomatic definition.

From here on, model will be used as a technical term for the equivalence class $[\Sigma]$ of Σ under the equivalence relation (2.6). With this terminology, the descriptive parameters relate to Σ and the intrinsic parameters relate to $[\Sigma]$. There is endemic confusion in econometrics between these two sharply different concepts, as will be discussed below (see, e. g., Section 3 (6)).

A similar distinction must be made also with regard to S. This is a rather subtle theoretical point. The intrinsic parameters of S arise from imposing certain restrictions on the data, such as the requirement that S admits a finite-dimensional realization. This will be discussed below and in Section 6.

The problem of model building, in the deterministic case, is to find a system Σ whose behavior is the same as the observed behavior $S_o = S_{\Sigma_o}$ of some (internally) unknown system Σ. If Σ is such a system, that is, if $S_\Sigma = S_o$, then we call Σ a realization of S_o. The mathematical problem is one of finding (F, G, H) given all the A_t in (2.5).

Evidently, the model $[\Sigma_{S_0}]$ serves as a kind of substitute for Σ_0. If we do not assume (and have reasons for assuming) that there is some (unknown) Σ_0 responsible for the generation of the (known) S_0, the whole modeling exercise becomes scientifically meaningless.

The abstract properties of a realization as just defined are sufficiently rich to allow important mathematical results to be obtained. The main fact is that for canonical realizations Σ^{can}, which always exist if any realization exists at all, there is a bijection (=one-to-one correspondence)

$$(2.7) \qquad S \xleftrightarrow{1\ -\ 1} [\Sigma_S^{can}]$$

between the data S and the model $[\Sigma_S^{can}]$.

Phrased differently, this is the classical result (1962) of deterministic realization theory that the data uniquely determines the model. For this reason, $[\Sigma_S^{can}]$ may be regarded as a very reasonable substitute for the (unknown) system Σ which generated the data S.

Thus we may always view the problem of model building as a deductive procedure described as

$$(2.8) \qquad \text{data} \longrightarrow \text{model},$$

where the mathematical specification of the arrow, and indeed the solution of the problem, is equivalent to the computation of the bijection claimed in Theorem (2.7).

The historical development of time-series analysis as well as of the related current econometric lore puts the cart before the horse. It attempts to determine the numbers specifying Σ_S from the numbers specifying S before the intrinsic (nonparametric) issues are understood.

By Theorem (2.7), the practical identification problem reduces to the determination of the intrinsic parameters of the model. Since there is a bijective correspondence between data and model, it is clear that the intrinsic parameters of the data (which is what we really mean by "data") must correspond bijectively to the parameters of the model. So the mathematical problem of intrinsic parametrization may be restated also as

(2.9) Parametrize data and model so that the bijective correspondence between the two is preserved.

The question of parameter identification doesn't even arise. Realization theory provides the rule for attaching, uniquely, the (a priori unknown) intrinsic parameters of $[\Sigma^{\text{can}}_{S}]$ to the (a priori known) intrinsic parameters of S. Every parameter of $[\Sigma]$ is "identifiable" because the correspondence (2.7) is bijective.

The (intrinsic) parametrization of $[\Sigma^{\text{can}}]$ is related to the old mathematical subject of canonical forms which is experiencing a renaissance under the impetus of system theory (see TANNENBAUM [1981]). As far as identification is concerned, the basic result is (2.7), which automatically proves that the "identifiable parameters" of a system based on the behavior data S are simply the intrinsic parameters of $[\Sigma^{\text{can}}_{S}]$. If we start from a given system Σ_o, then its intrinsic parameters are those of $[\Sigma^{\text{can}}_{S_{\Sigma_o}}]$.

In the applied context, the elements of the matrices A_1, A_2, ... are often called "behavioral parameters". This is an acceptably plastic terminology if we remember again that these are descriptive parameters, which have nothing to do with identification or identifiability. The are not the same as the intrinsic parameters. In fact, it is incorrect to assume that the elements of the sequence A_1, A_2, ... are arbitrary numbers because this may be in contradiction to the problem statement that S is to be realized by some (further specified or restricted) family if systems Σ.

The main object of the applied mathematical part of realization theory is the determination of the explicit numerical form of the bijective correspondence (2.7). Once this is known, we have in principle a computer program for the identification of $[\Sigma]$ from S.

Precisely because the correspondence (2.5) is bijective, every system property stated in terms of Σ must have a unique counterpart as a data property expressed in terms of S. The most important question of this sort is: How can we express the finiteness of Σ in terms of S? The answer to this question implicitly defines the intrinsic parameters of S.

If we define $\dim \Sigma :=$ size of the square matrix F_Σ, then the result is

$$(2.10)\qquad \dim \Sigma^{\mathrm{can}} = \operatorname{rank} B_S = \operatorname{rank} \begin{bmatrix} A_1 & A_2 & A_3 & \cdots \\ A_2 & A_3 & \cdots & \\ A_3 & \cdots & \cdots & \\ \cdots & & & \end{bmatrix}.$$

The condition imposed on B_S (sometimes called the *behavior* or *Hankel* matrix associated with the data S) shows that the intrinsic parameters of S are not free (unlike the descriptive parameters) but must satisfy the condition $\operatorname{rank} B_S = n$. The value of n is not known *a priori* (except that it must be finite) and is to be determined from the data S. Theorum (2.10) shows that this can be done, in principle, by computing the rank of the infinite behavior matrix B_S.

The preceding discussion outlines the questions that must be understood in order for the modeling problem to be *well-defined*. This is unfortunately not the case in a large part of the econometrics literature, as we shall show it in some detail in Sections 3-5 below.

In contrast to the deterministic case, the stochastic realization problem has not yet given rise to a definitive theory. The reasons for this lag may well lie in the preceding remark. In any case, as an organizing principle for wading through the existing conceptual mess, we adapt the nondebatable criterion

(2.11) *Any stochastic identification procedure must be effective also whenever the noise effects are arbitrarily small.*

Accordingly, it makes good theoretical sense to dissect the literature concerned with (linear) stochastic realization with respect to its treatment of the (linear) deterministic model which underlies *any* stochastic model.

Here our treatment of the stochastic aspects will be necessarily rather sketchy. This, however, is not as much of a limitation as it might seem.

The classical guiding principle of the stochastic analysis of linear systems is that all random inputs must be reduced, by suitable dynamical modeling, to white noise. (This principle is one of the main factors responsible for the success of Kalman filtering, as discussed in KALMAN [1960], [1978].) Thus *deterministic* dynamical modeling is almost always the main task confronting also stochastic realization theory; the analysis of the effects of white noise is straightforward and subsidiary. Thus

(2.12) *The first basic problem in* (linear) *time-series modeling is the precise and proper specification of the underlying* (linear) *deterministic dynamical system.*

3. IDENTIFIABILITY: FIRST EXAMPLE

To give concrete form to the preceding conceptual discussion, we may take, at random, some published material from the econometric literature dealing with identification and subject it to system-theoretic scrutiny.

Let us choose, for example, an expository article on identification by SCHÖNFELD [1979]. To facilitate referencing statements made in that article we reproduce here, in somewhat paraphrased English translation, the contents of Section 1 of that article. Italics are those of the original.

> "1. *Introduction.*
>
> "1.1 *Intuitive background.* The following discussion is restricted to the identification of *characteristics* [*parameters*] in stochastic models. In general terms, a characteristic is *identifiable* provided it can be uniquely inferred from the probability distributions of the observed random variables. In econometrics *identifiability* is to be regarded primarily as a necessary condition for the estimation of parameter one is interested in.
>
> "Let us now illustrate the intuitive basis of these definitions by an example."

At this point there is no possible objection to SCHÖNFELD's argumentation since it takes place on the level of intuitive "definitions". The difficulty immediately arises, however, that the notions of a *model* and *observed random variables* cannot be understood with sufficient precision on this (intuitive) level. This is seen by proceeding to dissect the example given by SCHOENFELD:

"1.2. Example. Consider the model with the equation

(a) $y_t = \alpha y_{t-1} + u_t, \quad t \in T = \{\ldots, -1, 0, 1, \ldots\},$

where $u = \{u_t\}$ is a linear stochastic process whose values

(b) $u_t = \sum_{\tau=0}^{\infty} \kappa_\tau \epsilon_{t-\tau}$

are generated with the aid of white noise $\epsilon = \{\epsilon_t\}$ defined by the assumptions

(c) $$\begin{cases} E(\epsilon_t) = 0, \quad E(\epsilon_t^2) = \sigma^2, \quad 0 < \sigma^2 < \infty, \\ E(\epsilon_t \epsilon_{t'}) = 0 \quad \text{for} \quad t \neq t', \quad t, \quad t' \in T, \end{cases}$$

where the nonstochastic parameters α, $\kappa = \{\kappa_t\}$ satisfy the conditions

(d) $|\alpha| < 1,$

(e) $\kappa_0 = 1,$

(f) $\Sigma_\tau |\kappa_\tau| < \infty.$

"Only the (wide-sense) stationary process $y = \{y_t\}$ is assumed to be observable."

"Conclusions:

(A) Under the general specifications of the model as given above, the parameter α is not identifiable. Reason: for each admissible "structure" $s' = (\alpha', \kappa', \epsilon)$ and each α'' there is a $\kappa'' = \{\kappa''_t\}$ such that y' and y'' from s' and $s'' = (\alpha'', \kappa'', \epsilon)$ are identical. The observed process y does not permit differentiating between various (assumed) values of α' and α''."

(B) If we "sharpen" the model in such a way that we "allow" only processes of moving average (MA) type to generate u, with the order of the MA process being no greater than L, in other words, if we define

(g) $u_t = \sum_{\tau=0}^{L} \kappa_\tau \epsilon_{t-\tau},$

then α is identifiable, with the exception of the "structures"

$Q_K(\alpha) = 0$ where

(b) $Q_K(z) = \Sigma_\tau K_\tau z^{L-\tau}$

(C) As a special case of (B), α will be identifiable if u is white noise ($L = 0$)."

"1.3. Remarks. (A) The identification problem occurs already with single equations.

(B) The same problem arises also with the parameters of a "reduced form".

(C) A parameter which is not identifiable in a model may very well become identifiable in a "sharpening" of this model. Identifiability depends crucially on the modeling assumptions, which therefore should always be given completely."

The analysis of the preceding assertions by SCHÖNFELD, which are typical of similar statements found in the literature, requires a large number of remarks.

(1) Note that "model" is not precisely defined. SCHÖENFELD implies that equation (a) is the "model". This is not enough; in addition to the state-transition equation given by (a), it is necessary to have a definition of input and output. The input should be defined as u_t given by (b-c) and the output as y_t. Here, accidentally, "state" and "output" are the same.

(2) The conditions imposed on u_t force the input to be the stochastic part of the problem. What is intended, evidently, is to define a deterministic (nonstochastic) model, namely equation (a) plus input plus output, which is to be identified from the probability distribution of y_t given only a-priori postulated stochastic properties (no observation) of the input.

Specifically, it is assumed that the input is generated by white noise acting as input on a linear, stable dynamical system (which is what the author means by the nonstandard term of linear process).

(3) For the deterministic identification problem involving (a) to make sense, the input would have to be known. Then it would follow (but see

below) that α is "identifiable". In formulating the problem in this way the basic question would be: Is the observed data y_t explainable by a *one*-dimensional model? In SCHÖNFELD's example this question is circumvented by the brute-force assumption that the system is, in fact, one-dimensional. This imposes a very strong prejudice on the data y_t.

(4) On the other hand, considering a random sequence u_t and postulating that it is generated by a linear system (b) with a (scalar) white noise input E_t is a very weak assumption. SCHÖNFELD alludes to this fact, rather imprecisely, by talking about weak stationarity. Roughly speaking, any weakly stationary process may be modeled in the manner described by (b) and (c). Thus weak stationarity is really the main assumption and not formula (b).

(5) Putting together the previous two remarks about y_t and u_t, we see that the example consists of the combination of a strong assumption (one-dimensionality) about the "model" for y_t with a weak assumption (weak-stationarity) about the "model" for u_t. This is, of course, intuitive nonsense.

(6) In precise terms, the specification of the example amounts to saying that (i) y_t is a weakly stationary process and (ii) the linear system generating y_t has the special property that it admits in its transfer function the factor $1/(z - \alpha)$. In other words, the hypothesis is that the system has a pole and the problem is to locate this pole by determining α.

(7) Now let us suppose that y_t (and hence, a *forteriori*, u_t) is generated by a finite-dimensional linear system Σ subject to white-noise input. In this setting, the problem posed by SCHÖNFELD is nonsense. Of course, every such finite-dimensional system Σ has a factor $1/(z - \alpha)$ in its transfer function. Every such factor is identifiable if and only if Σ, which is responsible for generating y_t *from white noise*, is identifiable. However, unless Σ = one-dimensional (which would be equivalent to SCHÖNFELD's very strong assumption (c)), *the problem of "identifying* α" *is not well defined* because the transfer function of Σ will have many α's and it is not clear *which* α is to be identified.

"Not well defined" necessarily implies "not identifiable". The real trouble is that the natural problem underlying the one circumscribed by SCHÖNFELD is not the identification of α but the identification of (the transfer function of) Σ, which is something entirely different.

(8) Next, taking the case complementary to (7), let us suppose that y_t is generated by an infinite-dimensional system. Then it is (at least in the elementary theory, under the assumptions stated by SCHÖNFELD) not clear, for reasons of mathematical rigor, exactly what is to be meant by a pole of a transfer function. (For example, the transfer function e^{-z} has no pole.) But the problem calls for locating the pole. Again the problem is not well defined.

(9) The obvious system-theoretic objection to SCHÖNFELD's example is that he arbitrarily singles out, by an ad-hoc assumption (forcing u_t to be nonobservable), a nonintrinsic property of the linear system Σ generating y_t. As we have seen in Section 2, the confusion arises from two common interpretations of the word "parameter", namely,

(i) descriptive parameters, and

(ii) intrinsic parameters.

SCHÖNFELD uses "parameter" in the first sense when he writes down equation (a). When he talks about "identifying α", SCHÖNFELD reverts to the second sense. Evidently α is not an intrinsic parameter of Σ and so we cannot talk about "identifying" it.

In other words, since equation (a) is not a proper way of specifying the model relevant to the example, the parameter α that went into (a) cannot be recovered from the data constituted by the probability distribution of y_t. Only intrinsic system properties can be expected to be "identifiable", not things like the choice of a coordinate system, choice of units of measurement, etc.

Clearly the problem is not well-defined.

(10) When SCHÖNFELD speaks, under 1.1, of the "estimation of parameters one is interested in" he makes it quite clear that he regards a "parameter" as an absolute attribute of a system. Apparently he takes it for granted

that there exist such "absolute" parameters.

Unfortunately, this is merely wishful thinking and not a theorem. In general, mathematical objects, such as a linear system, cannot be parametrized in such a way that parameters have absolute significance like, for example, mass in physics. Mass possesses this desirable attribute because it is directly observable and context-independent. Economic quantities (like inflation, rate of savings, productivity, etc., etc.) have no such absolutely measurable attributes but are very much context-dependent and interrelated with many other economic variables. For example, inflation does not mean the same thing in a classical economy as in a socialist one or in the hypothetical one where all wages, taxes, savings, etc. are perfectly indexed.

Thus what we may hope to identify in an interrelated situation is a model but not a specific and rigidly given system of coordinates (intrinsic parameters) for that model.

(11) There is a further element of fuzziness in SCHÖNFELD's description of the problem. He does not specify with mathematical precision what is to be regarded as data for the identification problem. (He mentions under 1.1 that identification is to be based on the "probability distributions of the observed random variables". This is not enough if we want to do calculations; for them, further information or assumptions are needed about the structure of the underlying probability distributions.)

The conventional formulation is as follows. We take y_t (or u_t) as a Gaussian (or, equivalently, from the theoretical point of view, second-order) random sequence. Then the problem data consists of the knowledge of the covariances

([illegible].1a) $\operatorname{cov}(y_t, y_{t+r}), \quad r = 0, 1, 2, \ldots ,$

in addition to the relatively harmless normalizing assumption

([illegible].1b) $y_t \quad 0.$

With these specifications, we finally have a well-defined problem. It is to determine the equivalence class $[\Sigma]$ of all linear systems Σ which

COULD HAVE generated the data (3.1). An element Σ of $[\Sigma]$ is then called a realization of (3.1) and $[\Sigma]$ is the model or models we wish to identify.

In practice, we pick out a "typical element" Σ_* of $[\Sigma]$ and identify that. If the realization problem fails to have a unique solution, then, unfortunately, the class $[\Sigma]$ will contain more than one essentially different model; to classify the family of essentially different elements of $[\Sigma]$ we will need a new kind of "parameters" which, by definition, cannot be obtained from the date (3.1).

(12) What we have just described is an example of the stochastic realization problem in system theory. This problem has an enormous literature (see, e.g. KALMAN [1965], RISSANEN and KAILATH [1972], FAURRE [1973], AKAIKE [1974], [illegible] [1976, 1977], FAURRE et al. [1977], VAN PUTTEN and VAN SCHUPPEN [1979]). The problem is not yet completely settled today in that the precise determination of the equivalence class of all realizations of stochastic data like (3.1) is mathematically nontrivial.

(14) Thus we have arrived at a reformulation of the problem which differs substantially from the point of view taken by SCHÖNFELD.

In accordance with the prescription

$$\text{data} \longrightarrow \text{model},$$

the problem is to determine first (usually only abstractly) the class $[\Sigma]$ of all models Σ possessing the behavior fixed by the data (3.1). For example, the problem may be posed in such a way that Σ must be a linear model, which is the counterpart of the assumption that the data is of the form (3.1).

Having determined the class $[\Sigma]$, there are two remaining theoretical questions:

Uniqueness. Does $[\Sigma]$ have more than one essentially different element? (The next section offers an example of this.)

Parametrization. Parametrize the family of all possible data (3.1), thereby automatically parametrizing the family of all equivalence classes $[\Sigma]$. This is usually a deep mathematical problem in the general realm of system theory which in many cases, especially those of interest to econometrics, is presently open. (Of course, parametrization, in the mathematical sense, is always what we have called intrinsic parametrization.)

Thus the question of "identifiability of parameter" does not arise at all. "Parametrizing the data" means that the data parameters (which correspond to the second type under (9)) are identifiable by definition. The family of all $[\Sigma]$ models is then automatically parametrized because its elements correspond bijectively to the problem data.

(14) The preceding may become clearer if we now make the obvious remark that a linear system (like (a)) driven by another linear system (like (b)) subject to a white noise input is still simply a linear system subject to a white noise input. If Σ_1 is this linear system, it may be factored (cascaded) as

$$(5.2)\qquad \Sigma_1 = \Sigma_2\Sigma_3,$$

with Σ_2 generating y_t from an unobserved stochastic input u_t and Σ_3 generating u_t from an assumed (but again unobserved) white noise sequence ϵ_t. Such a factorization is not an intrinsic property of the system Σ_1. It can be performed in many ways. There is no reason for doing it so that Σ_2 = 1-dimensional (the SCHÖNFELD assumption). Since the factorization is necessarily commutative, no useful statements can be made about the dimension of either factor except that $\dim \Sigma_1 = \dim \Sigma_2 + \dim \Sigma_3$. So we see once again that the SCHÖNFELD problem is not well defined.

(15) When SCHÖNFELD goes on to talk in Section (B) about "sharpening" the model, the system theorist objects. To say that u_t is generated by a moving average process is an assumption completely out of the blue, unless it is known (for some special extraneous reason) that this is so, in which case the definition of the model must be modified to give a new "parameter set" $(\alpha, Q(z))$. See (b).

(16) To be well-defined, this new parameter set must be subjected to

the restriction that $Q(z)/(z - a)$ has no common factor (i.e., that $Q(\alpha) \neq 0$). The "exceptional case" $Q(\alpha) = 0$, which SCHÖNFELD interprets as destroying the identifiability of α, is not a case at all as it must be ruled out in advance for the problem to be well defined. So $Q(\alpha) \neq 0$ has nothing to do with identifiability.

(17) If $\{(\alpha, Q(z)), Q(\alpha) \neq 0\}$ is taken as the new parameter set, corresponding restrictions must be imposed also on the probability distributions of the observed random variables. This means that the data (3.1) must now be restricted by the condition that it is generated from white noise via the model with transfer function $Q(z)/(z - \alpha)$, cancellation not allowed. It is a serious shortcoming of SCHÖNFELD's discussion that he overlooks this point.

(18) SCHÖNFELD's further "sharpening" of the model, namely the postulate that u_t is a white-noise sequence leads to the correct claim that α is identifiable, because then Σ, the linear system which generates y_t from white noise, is one-dimensional and questions of factorization and parametrization are trivial. Far from being a "sharpening" of the model, this is in fact a very strongly prejudiced assumption. To fix the dimension of Σ to be 1 irrespective of the data does violence to the problem of identification since, as we have argued before, dim Σ is information that must be deduced from the data and not imposed beforehand.

(19) SCHÖNFELD's explanatory remarks under (1.3) are now irrelevant. In particular, it is not true that "a parameter which is not identifiable in a model may well become identifiable in a sharpening of this model". What has happened is that α was not well-defined initially as an intrinsic parameter because the model was not well-defined; if after imposing very strong restrictions the model becomes well-defined it might well also happen that the parameter becomes intrinsic. This is not an illustration of identifiability but a symptom of a bad theory.

We may roughly summarize these critical comments by noting the following basic facts of life in modeling and identification:

(i) The first theoretical task in any identification problem is to make sure the model (system to be identified) is well defined.

(ii) Given (i), it is possible to compute (in principle) the equivalence class $[\Sigma]$ of all systems Σ which generate the same (external) data.

(iii) By Theorem (2.7) data parametrization induces a parametrization of the family of all equivalence classes $[\Sigma]$. By definition, we are dealing here with intrinsic parameters and the question of "parameter identifiability" is empty.

(iv) If the realization problem has a unique solution, we are finished.

(v) If the identification problem has a nonunique solution, then further parameters may have to be introduced to describe all the models in $[\Sigma]$ which are essentially different from each other; such parameters are of course never identifiable.

To put the issue even more crudely and briefly, parameter indentifiability is not a viable scientific concept. The real problems concern the uniqueness of realization and applied mathematical techniques for computing realization.

We do not wish to create the impression that SCHÖNFELD's article was quoted out of context. Very similar remarks apply, for example, to the introductory discussion of HANNAN [1971]. The examples given by him illustrate situations where the model is not well defined. The source of confusion is again the failure to make distinctions between descriptive and intrinsic parameters. Further examples may be found in KALMAN [1981].

4. IDENTIFIABILITY: SECOND EXAMPLE*

To illustrate further the notion of a stochastic realization we shall now look at an example taken from the classical paper of KOOPMANS[†] and REIERSØL [1950]. It is concerned with the identification of a static relationship and in that sense it is trivial from the point of view of (dyna-

---

[*] This section was not part of the oral presentation.

[†] I am indebted to Professor Koopmans for having pointed out to me this and other related papers, nearly fifteen years ago already.

mical) system theory.

Consider a linear (more precisely, affine) relationship between a scalar *input* u and a scalar *output* y given (necessarily) by two unknown parameters α, β:

$$(4.1)\qquad y = \alpha + \beta u.$$

The information obtained about this relationship comes from two noisy observations, which we shall write (in accordance with the classical notations of "Kalman filtering theory" in KALMAN [1960] as

$$(4.2)\qquad \begin{cases} z_1 = u + v_1, \\ z_2 = y + v_2, \end{cases}$$

with v_1, v_2 random variables subject to the assumption

$$(4.3)\qquad E(v_i) = 0, \quad i = 1, 2.$$

The *model* for this problem is given by (4.1-2) and by the preceding specification of input, output, and observations. (In general, and also here, for stochastic systems the output is not necessarily the same as the observables.) The *stochastic environment* is specified by (4.3) and by some assumption concerning the probabilistic relationships between u, v_1, and v_2. With KOOPMANS and REIERSØL we assume that u is gaussian and independent of v_1 and v_2.

Having set the stage, the problem is now: *Are* α, β *identifiable or not identifiable* (under various specific assumptions concerning the joint distribution of (v_1, v_2), in addition to those just stated above)?

Case (A). The noise (v_1, v_2) is jointly gaussian. Under the assumptions stated, we find that

$$(4.4)\qquad E(z) = \begin{bmatrix} E(u) \\ E(y) \end{bmatrix} = \begin{bmatrix} E(u) \\ \alpha + \beta E(u) \end{bmatrix} .$$

and

(4.5) $$\operatorname{cov} z = \begin{bmatrix} 1 & \beta \\ \beta & \beta^2 \end{bmatrix} \operatorname{var} u + \operatorname{cov} v.$$

Recalling also the zero-mean assumptions, it follows that the stochastic environment is specified by five unknown parameters, namely

(4.6) $$E(u), \quad a = \operatorname{var} u, \quad b = \operatorname{var}(v_1^2), \quad c = \operatorname{cov}(v_1 v_2), \quad d = \operatorname{var}(v_2^2),$$

subject to the restrictions

(4.7) $$a \geqq 0, \quad \operatorname{cov} v = \begin{bmatrix} b & c \\ c & d \end{bmatrix} \geqq 0.$$

Under these hypotheses, the probability distribution of the vector z is gaussian and is specified by $E(z)$ and $\operatorname{cov} z$. Making the conventional assumption that $E(z)$ and $\operatorname{cov} z$ may be determined or estimated (since the vector z is observable), KOOPMANS and REIERSØL conclude that the pair (α, β) is not identifiable on the basis of the data $E(z)$ and $\operatorname{cov} z$.

Their reasoning is as follows. Assuming $\operatorname{cov} z$ is nonnegative definite (as it must be for a covariance), there are in general many parameters β which satisfy the obvious constraint

(4.8) $$\begin{bmatrix} 1 & \beta \\ \beta & \beta^2 \end{bmatrix} a \leqq \operatorname{cov} z.$$

For each such β the condition (4.5) can be met by suitable choice of the environmental parameters $a, \ldots, d$ satisfying the constraints (4.7). For each β equation (4.4) defines a unique value of α.

Even allowing for the embryonic state of system theory in 1950, I would doubt if any system theorist would have accepted this conclusion of KOOPMANS and REIERSØL in the manner in which it was first stated. The difficulty revolves around the proper definition of the model and of its stochastic environment.

First, let us assume that $c = \operatorname{cov}(v_1 v_2) = 0$, in other words, that v_1 and v_2 are independent. Then the identification equations become

$$(4.9)\quad \begin{cases} \operatorname{cov}(z_1z_2) = \beta a, \\ \operatorname{var}(z_1^2) = a + b, \\ \operatorname{var}(z_2^2) = \beta^2 a + d \\ E(z_1) = E(u), \\ E(z_2) = E(y) = \alpha + \beta E(u). \end{cases}$$

The same setup is used in MEHRA [1976, p. 192-193].

A solution (β, a) of the first three equations, if it exists, must satisfy the conditions

$$(4.10a)\quad \beta a = \operatorname{cov}(z_1z_2),$$

$$(4.10b)\quad \operatorname{var} z_1 \geqq a,$$

$$(4.10c)\quad \frac{\operatorname{var} z_2}{\overline{\operatorname{cov}(z_1z_2)}} \geqq |\beta|.$$

By $\operatorname{cov} z \geqq 0$ a solution satisfying these relations always exists; in fact, the set of such (β, a) is just the segment of the hyperbola (4.10a) delimited by the inequalities (4.10b) and (4.10c). (In other words, we have here an elementary algebraic-geometric problem calling for locating a certain segment of a curve given by algebraic equations and inequalities: the existence of the solution coincides with the probabilistic requirement that $\operatorname{cov} z \geqq 0$). Given any such admissible pair (β, a), the values of b and d are determined from (4.5) and α is determined from (4.4). Obviously the solution is not unique.

Second, the assumption that $\operatorname{cov}(v_1v_2) = 0$ is not a luxury or a loss of generality but indispensable for the correct discussion of this problem. System-theoretically the problem is to estimate the (affine) effect of a variable u on another variable y when u and y can be observed only in a noisy way as z_1 and z_2. Recall that the only assumption (knowledge) about v_1 and v_2 is (4.3) and that they are jointly gaussian. If there were (nonzero) correlation between the values of v_1 and v_2 this would imply that part of the cause-and-effect relationship between z_1 and z_2 is to be explained by the model (4.1) whereas another part, of unknown amount, is to be explained by the correlation between the noise components about which nothing is known quantitatively. (Indeed, if $\operatorname{cov}(v_1v_2) \neq 0$ we may always regard v_2 as generated by another linear model $v_2 = \hat{\beta}v_1 + \hat{v}_2$, with $\hat{\beta}$ fixed by the covariance between v_1 and v_2 and $\hat{v}_2$ independent of v_1.) This is a highly unnatural assumption if the exercise is to be relevant to the real world.[*] Hence in this respect the KOOPMANS-REIERSØL example must be modified by insisting on $\operatorname{cov}(v_1v_2) = 0$.

The nonuniqueness of the solution of this realization problem is unfortunately a fact of life and has been observed in many contexts. (The interesting discussion of WOLD [1972] concerning the need for causality assumptions in addition to statistical techniques in the treatment of regression is highly enlightening here.)

The parameter α plays an uninteresting role in the preceding analysis since the treatment of the mean is essentially a deterministic linear system problem. Notice also that the assumption $E(v_1) = E(v_2) = 0$, which is made by KOOPMANS and REIERSØL, plays a role similar to the assumption $\operatorname{cov}(v_1v_2) = 0$, which was not made by them. If we didn't assume $E(v_1) = E(v_2) = 0$ a part of α would be explained by the model (4.1) and another part by the noise.

Case (B). The noise (v_1, v_2) is not gaussian. In this case, it can be shown that (α, β) are identifiable. (For the general analysis of the problem, see REIERSØL [1950].) System-theoretically this statement, too, requires discussion. If the joint distribution of (z_1, z_2) is not gaussian then, roughly speaking, the cause-and-effect relation between them is not linear. Consequently this result of REIERSØL must be viewed as an isolated glimpse at nonlinear realization theory. Having made such an odd assumption about the noise it is mandatory to explain why the implicit nonlinear dependence between z_1 and z_2 should be modeled linearly by (4.1).

[*] To the physicist, the KOOPMANS-REIERSØL example amounts to having to estimate the values of two resistors connected in series such that only the sum of resistances is available for measurement.

Since nonlinear realization theory is not yet a developed subject, it is not possible at present to give a deductive discussion here in the style of Case (A).

In summary, we may say:

(i) The KOOPMANS-REIERSØL example is a system-theoretic problem.

(ii) It is not well defined unless $c = \text{cov}\ (v_1 v_2) = 0$, for otherwise the causal effects are arbitrarily divided between a model to be determined and noise about which nothing is assumed or observable.

(iii) With the normality assumptions and $c = 0$, we have a well-defined (elementary) problem in linear stochastic realization theory.

(iv) This problem, unfortunately, has a nonunique solution, which is typical of stochastic realization problems. This conclusion agrees with the intuitive judgment that such formulations are acceptable "provided the temptation to specify models in such a way as to produce identifiability of relevant characteristics [i.e., parameters] is resisted" (from KOOPMANS and REIERSØL [1960, Sect. 2.4, p. 169]).

(v) REIERSØL's analysis of the case $z \neq$ gaussian will be a benchmark of nonlinear realization theory when the latter comes into being.

(vi) MEHRA [1976] shows that the nonuniqueness of the stochastic realization problem posed by KOOPMANS and REIERSØL (with $\text{cov}\ (v_1 v_2) = 0$ being explicitly assumed by MEHRA) can be removed by reformulating the problem. MEHRA assumes that both u and v_1 are random sequences, the first having correlation properties and the second being white. Then it is possible to determine $E(u^2)$ and $E(v_1^2)$ separately, not just their sum $E(z_1^2)$, which makes the solution unique.

This is a much more realistic formulation of the problem; it recognizes the distinction between the causal variable (the sequence u) and the noise (the sequence v_1), while KOOPMANS and REIERSØL grant u and v_1 substitute for causality hypotheses in any statistical identification problem since statistics is neutral about what causes what.

5. THE "ARMA" MODEL

The widespread use in econometrics of ARMA models (see BOX and JENKINS [1976]) raises system-theoretic problems which require discussion. We can now amplify the comments made in Section 3.

The general model of this type is often described in econometrics in the following terms (see, for example, DEISTLER [1978]):

$$(5.1) \qquad \sum_{r=o}^{n_1} Q_r y_{t-r} = \sum_{s=o}^{n_2} N_s u_{t-s} + v_t.$$

The vector variables y_t and u_t have the same intuitive significance as in the example of Section 3; the vectors v_t are additional error terms.

We shall consider only the deterministic aspects of the problem. They revolve around the question: In what sense does (5.1) determine a linear *system*?

(1) Equations (5.1) describe a system in the external sense; there are no state variables.

(2) For the output sequence y_t to be determined from in input sequence u_t we must have

$$(5.2) \qquad \det Q(z) \neq 0$$

(as a polynomial), where $Q(z)$ is the matrix polynomial

$$(5.3) \qquad Q(z) := \sum_{r=0}^{n_1} Q_r z^{n_1-r}.$$

(DEISTLER assumes $\det Q_o \neq 0$; this is unnecessarily restrictive.)

(3) Now we can write $Q^{-1}(z)N(z)$, with $N(z)$ the matrix polynomial

$$(5.4) \qquad N(z) := \sum_{s=o}^{n_2} N_s z^{n_2-s}.$$

$Q^{-1}(z)N(z)$ is a rational matrix and therefore it has a formal Laurent series about $z = \infty$. To be able to relate this series to the transfer function of the underlying system Σ, it is necessary, for reasons of causality and normalization, that

(5.5) $Q^{-1}(z)N(z)$ – proper rational matrix.

(DEISTLER does not discuss this point.)

(4) The last assumption means that

(5.6) $Q^{-1}(z)N(z) = \sum_{t>0} A_t z^{-t}$,

Where = holds in the sense of formal power series. We are now in the situation of having given a standard external description $S = (A_1, A_2, \ldots)$ of the system in the form (2.4). Of course, S is "identifiable", by definition, since this is the data to which the identification problem is ultimately referenced.

(5) DEISTLER refers to "conditions" for the identifiability of (5.5). Presumably he means thereby relations between the descriptive parameters of $Q(z)$, $N(z)$, which are the matrices $Q_0, Q_1, \ldots, Q_{n_1}$; $N_0, \ldots N_{n_2}$, and the descriptive parameters of S, which are the matrices $A_1, A_2, \ldots$. However, one cannot consider such things until after the system is well defined. This obviously requires introducing the equivalence relation

(5.7) $(Q(z), N(z)) \sim (\hat{Q}(z), \hat{N}(z))$

defined by

(5.8) $Q^{-1}(z)N(z) = \hat{Q}^{-1}(z)\hat{N}(z)$.

The equivalence relation (5.7) is needed to prevent the system to be ill-defined due to the possibility of cancelable factors between $Q(z)$ and $N(z)$.

(6) Conditions (5.2), (5.5), and the equivalence relation (5.7) are obviously necessary for the abstract map

(5.9) $(Q(z), N(z)) \longrightarrow S$

given by (5.6) to be injective.

To establish that this map is bijective---in other words, that we can legitimately talk about (5.1) as a well-defined external description of some underlying system---we must show that there exists an injective map

(5.10) $S \longrightarrow (Q(z), N(z))$

satisfying (5.6) This is mathematically nontrivial.

Assuming S has a finite-dimensional realization, realization theory shows that there is a proper rational matrix $Z(z)$ whose formal power series agrees with the right-hand side of (5.6). It can be shown further that every proper rational matrix admits a factorization as $Z(z) = Q^{-1}(z)N(z)$.

This proves the existence of an injective map (5.10). Then it follows that the correspondence between S and $(Q(z), N(z))$ can be made bijective with the aid of the conditions mentioned above.

(7) The bijective relationship between the two external descriptions is established abstractly and has nothing to say about the descriptive parametrizations mentioned under (4). The question of intrinsic parametrization is left open. For S, the dimension of the corresponding canonical realization is given by (2.10). It is necessary to prove an analogous formula for $(Q(z), N(z))$. Then canonical forms can be derived for $Q(z)$ and $N(z)$ which exhibit the intrinsic parametrization. This is by no means a simple matter mathematically; the reader is referred to KALMAN [1981].

In any case, it is certainly not possible to fix the values n_1 and n_2 (as is done by DEISTLER [1978]) because these quantities do not have any simple relationship to the underlying system Σ. In fact, it is not possible to give a single, globally valid formula for the object $(Q(z), N(z))$ because its intrinsic (canonical) parametrization requires knowledge of the so-called Kronecker indices (see KALMAN [1971]. These are a global property of the underlying Σ which govern the places where the canonical parameters appear in the matrices $Q_o, \ldots N_{n_2}$.

Subject to these technical details, however, the question of how to define a deterministic linear dynamical model (external sense) with the aid of a pair $(Q(z), N(z))$ is completely settled at the present time. (This amounts to assuming u_t = known, $v_t = 0$.) It should be mentioned that the mathematical analysis required here is relatively recent even in system theory and was developed mainly during the last ten years under the impetus of the book of ROSENBROCK [1970].

(8) When we come to the stochastic case (the probability distributions of u_t, v_t are known but not the actual values), then it would seem (to the writer) that the situation is not yet clear, in spite of much recent work. In particular, it is not clear under what general conditions, if any, the stochastic realization problem admits a unique solution.

This discussion again shows that "parameter identifiability" is a nonproblem for the model formalized in terms of (5.1).

The ARMA scheme defines a generic linear system. This is undoubtedly a major reason for the success of methods based on it. Dropping either the "AR" or the "MA", in other words, considering only moving-average or autoregressive models, genericity is lost and further serious conceptual difficulties arise.

6. MORE PITFALLS

"Parameter identifiability" is, intuitively, a rather appealing notion. Why did it fail for linear systems? The reason is simply that the development of a particular field of science---system theory---has reached a point where results from a subfield---realization theory---are available to subject the intuitive notion of "parameter identifiability" to rigorous scientific analysis. When this is done, "parameter identifiability", after having been tested on the precise and concrete case of a linear system, collapses as a workable concept---and yet it must work for linear systems if it has any theoretic merit at all.

A similar application of system theory can be made to assess the merits of the moving average (MA) and autoregressive (AR) types of models, which were proposed by time-series analysis long before system theory existed (certainly by the late 1920's). The study of these very models provided some of the stimuli for the development of system theory.

Are these models good or bad? Any good system theorist would immediately reply, "bad". The justification of this emotional conclusion by rigorous methods, however, is not at all trivial; in fact, it was made possible only by the recent development of the so-called "partial realization" theory (see KALMAN [1971, 1979c]).

The partial realization problem arises when S in (2.4) is given only partially, that is, by a finite sequence of matrices $A_1, \ldots, A_t$. Then the realization $\Sigma^{\min}$ of minimal dimension n_t may not be unique, but of course n_t is unique because of the requirment of minimality. The analysis of n_t as a function of t *provides very important mathematical information* concerning the classical realization problem (see KALMAN [1979c]). Since n_t is a monotone, nondecreasing, integer-valued function on the integers, its value can change only in "jumps". The structure of these jumps turns out to satisfy strong regularity conditions (see below). The generic case occurs when all jumps are equal to one and the jumps occur at all odd values of T.

The AR and the MA schemes constitute a constructive existence proof that for every t the partial realization problem has a (finite) solution. Mathematically, this is a very trivial fact. It follows, unfortunately, that this is the only system-theoretic idea inherent in AR and MA.

To be precise, the AR scheme applies if and only if n_t is a function with a single jump. This is a highly nongeneric situation. Consequently it usually does not occur in nature. Since the "jumps" are the basic phenomena in the realization problem, it must be possible to find a statistical test for the occurrence of various patterns of jumps.

Such a test has not yet been developed (in the writer's knowledge).

Consequently, the application of AR models to real data is system-theoretic nonsense. When an AR scheme is applied, usually it is out of pure prejudice, without statistical evidence that the very unlikely case is in fact at hand. It cannot be argued that AR fits the data because ARMA will fit the data even better since ARMA is generic (see Section 5).

It is an amusing fact that for $n = 1$ it is trivially true that AR = ARMA. Consequently there is nothing wrong in applying first-order autoregression; it is when we attempt to jump to the n-th order case that the idea collapses. Generalizations of this type frequently lead to a system-theoretic problem which is far more difficult than it seems.

Similar comments apply of course also to the MA scheme.

Another application of partial realization theory concerns the "parameters" of the data $S = (A_1, A_2, \ldots)$. Let us take the scalar case $S = (a_1, a_2, \ldots)$, where the a_i are real numbers, since the theory given in KALMAN [1979c] is restricted to this case. Such a scalar sequence may arise, for example, as a discrete-time autocovariance function.

The theory in KALMAN [1979c] applies to any such sequence (there are no conditions!). Consequently any "data" $a_1, a_2, \ldots$ has a certain pattern of jumps associated with it. If a jump of size q_i occurs at the time point t_i, the following statements can be made:

(i) The value of $a_{t_i} \neq a^*_{t_i}$:= value of a_{t_i} computed from the (unique) partial realization based on $a_1, \ldots, a_{t_{i-1}}$, using formulas (2.5). (This statement is meaningful because the main theorem of partial realization theory guarantees that "there can be no jump before uniqueness".) Thus a_{t_i} is not an arbitrary parameter because otherwise there is a contridiction to the inherent jump pattern.

(ii) After a jump of size q_i exactly q_i elements $a_{t_i+1}, \ldots, a_{t_i+q_i}$ of the sequence are "free"; that is, any values of these parameters may occur without contradicting the jump pattern.

(iii) Before a jump of size q_i, exactly $q_i - 1$ elements of the sequence are fixed, that is, they are uniquely determined via (2.5) by the partial realization based on the first $2n_{t_{i-1}}$ elements of the sequence.

This shows that it is very naive to speak of $S = (a_1, a_2, \ldots)$ as a sequence of (intrinsic) parameters, if we take parameter in the usual sense of being any (real) number. Only the q_i elements of the sequence following the i-th jump point (type (ii)) qualify as parameters in this sense. The type (i) elements must satisfy an $\neq$ condition and type (iii) elements (which do not occur in the generic case) are completely fixed. Moreover, and this is the crucial point, the location of the various types of elements is fixed by the jump pattern, which, as an intrinsic property of the data, has not been discovered in the time-series literature prior to KALMAN [1971, 1979c].

Thus we see that at present even the preliminary question of parametrization of the data raises deepseated theoretical questions.

7. CONCLUSIONS

System theory is a new paradigm. It applies to economics in at least two ways: through the study of systemic properties of any economic model and through the critique of the econometric recipes of modeling. We have studied here the second aspect.

The development of science, certainly in regard to economics, has now reached the stage where Newton has become a very poor role model. The Newtonian approach---isolate a phenomenon and try to analyze it in its simplest appearance without regard to the context---is precisely what is inapplicable to the problems of economics, because economic phenomena are intrinsically system (context) related.

The aspiration of econometrics, to develop hidden quantitative relationship from interrelated data, is a problem which belongs to the domain of system theory. To be successful, this program must be supplemented by close attention to more advanced problems arising from modeling, for example, the question of intrinsic parameters and their determination from real data. It is certainly not enough to think in terms of parameters in the naive sense. In particular, we have shown that the intuitive notion of "identifiability of a parameter" cannot be developed into a meaningful scientific concept and must be replaced by the apparatus of realization theory.

Let us close by recalling the position taken twenty-five years ago by John von NEUMANN [1955] in one of his last public statements. In response to a discussion topic concerning the potential scientific development of economics, he denied that progress would be impeded because of the "impossibility of experiments" (classical astronomy, a successful science without experiments, being his counterexample) or hindered by "lack of data" (many scientific advances, such as EINSTEIN's photoelectric law and even more his general relativity, were conceived when there was very little available data). The most important missing element in economics, von NEUMANN insisted, was the "definition of categories".

What he meant by this is articulated in contemporary terminology by the words "invariants", "structure", "decomposition", "elements", etc. If modeling of economic time series is to have relevance for economic theory---and in that hope we are all united---then system theory must be able to dig up von NEUMANN's missing categories by penetrating more deeply into the theory of models.

There are many results and much research in system theory today concerned with exactly the same problem. Econometrics must also contribute to its solution, or wither as an irrelevant exercise in statistics.

The task of going from real data to economic (or any other) theory may be attacked in many ways. The scientific approach is not compulsory. Astrology has been tried. No optimist would quarrel with the declaration of one of von NEUMANN's direct successors that "exposure to the 'real world' of economic policy...affirmed rather than eroded [my] belief in the usefulness and relevance of economic theory" (see WHITMAN [1979, page x]). Innocent faith sometimes moves mountains. But don't hold your breath. It is better to sit down and to begin by cleansing the field of misconceptions.

H. AKAIKE

[1974] "Stochastic theory of minimal realization", IEEE Transactions on Automatic Control, **AC-19**: 667-674.

G. E. P. BOX and G. M. JENKINS

[1976] *Time Series Analysis: Forecasting and Control*, revised edition, Holden-Day, 575 pages.

M. DEISTLER

[1978] "The structural identifiability of linear models with autocorrelated errors in the case of cross-equation restrictions", Journal of Econometrics, **8**: 23-31.

P. FAURRE

[1973] "Réalisations markoviennes de processus stationnaires", Research Report No. 13, IRIA, Rocquencourt, FRANCE.

P. FAURRE, M. CLERGET, and F. GERMAIN

[1979] *Opérateurs Rationnels Positifs: Application á l'Hyperstabilité et aux Processus Aléatoires*, Dunod, 1979, 294 pages.

E. J. HANNAN

[1971] "The identification problem for multiple-equation systems with moving average errors", Econometrica, **39**: 751-765.

R. E. KALMAN

[1960] "A new approach to linear filtering and prediction problems", Journal of Basic Engineering (Trans. ASME), **82 D**: 35-45.

[1965] "Linear stochastic filtering theory---reappraisal and outlook", in *Proc. Symposium on System Theory*, edited by J. Fox, Polytechnic Institute of Brooklyn, pages 197-205.

[1969] *Topics in Mathematical System Theory*, McGraw Hill, 358 pages.

[1971] "On minimal partial realizations of a linear input/output map", in *Aspects of Network and System Theory* (a collection of papers in honor of E. A. Guillemin), edited by R. E. Kalman and N. DeClaris, Holt, Rinehart, and Winston, pages 385-408.

[1972] "Kronecker invariants and feedback", in *Proc. 1971 NRL-MRS Conference on Ordinary Differential Equations*, edited by L. Weiss, Academic Press, pages 459-471.

[1976] "Realization theory of linear dynamical systems", in Control Theory and Functional Analysis, Vol. II, International Atomic Energy Agency, Vienna, 1976, pages 235-256.

[1978] "A retrospective after twenty years: from the pure to the applied", in Applications of Kalman Filter to Hydrology, Hydraulics, and Water Resources, edited by Chao-Lin Chiu, Department of Civil Engineering, University of Pittsburgh, pages 53-69.

[1979a] "A system-theoretic critique of dynamic economic models", in Global and Large-Scale System Models, edited by B. Lazarevic, Springer, pages 1-24.

[1979b] "Theory of modeling", in Proceedings of the IBM System Science Symposium, Oiso, JAPAN, edited by Y. Nishikawa, pages 53-69.

[1979c] "On partial realizations, transfer functions, and canonical forms", in Acta Polytechnica Scandinavica, Mathematics and Computer Science Series No. 31, pages 9-32.

[1980a] "System-theoretic critique of dynamic economic models", International Journal of Policy Analysis and Information Systems, 4: 3-22.

[1980b] "Dynamic econometric models: a system-theoretic critique" in New Quantitative Techniques for Economic Analysis, edited by A. Cellina and G. P. Szegö, Academic Press, pages

[1981] "Finitary models for time series and the identification of dynamics", in Developments in Statistics (book), edited by P. R. Krishnaiah, Academic Press, pages

T. C. KOOPMANS and O. REIERSØL

[1950] "The identification of structural characteristics", Annals of Mathematical Statistics, 21: 165-181.

R. K. MEHRA

[1976] "Identification and estimation of the error-in-variables model (EVM) in structural form", in Mathematical Programming Study (North Holland) vol. 5, pages 191-210.

J. von NEUMANN

[1955] "The impact of recent developments in science on the economy and on economics", summary of speech before the National Planning Association, Washington, DC; in Looking Ahead, 4: 11; reprinted in the author's Collected Works, Vol. VI, pages 100-101.

G. PICCI

[1976] "Stochastic realization of gaussian processes", Proceedings of the IEEE, 64: 112-122.

[1977] "Some connections between the theory of sufficient statistics and the identifiability problem", SIAM Journal of Applied Mathematics, 33: 383-398.

O. REIERSØL

[1950] "Identifiability of a linear relation between variables which are subject to error", Econometrica, 18: 375-389.

J. RISSANEN and T. KAILATH

[1972] "Partial realization of random systems", Automatica, 8: 389-396.

H. H. ROSENBROCK

[1970] State-Space and Multivariable Theory, Wiley, 257 pages.

P. SCHÖNFELD

[1979] "Identification" (in German), in Encyclopedic Handbook of the Mathematical Economic Sciences, edited by M. Beckmann, G. Menges, and R. Selten, ... pages 61-66.

A. TANNENBAUM

[1981] Invariance and System Theory: Algebraic and Geometric Aspects, Springer Lecture Notes in Mathematics No. 999, pages.

C. VAN PUTTEN and J. H. VAN SCHUPPEN

[1979] "On stochastic dynamical systems", Proceedings 4th International Symposium on the Mathematical Theory of Networks and Systems, July 1979, Delft, NETHERLANDS]

M. v. N. WHITMAN

[1979] Reflections of Interdependence: Issues for Economic Theory and US Policy, University of Pittsburgh Press, 316 pages.

H. WOLD

[197[illegible]]

Synthèse,

DATE

ILME